Sampling representative years for a TSO in a climate simulation of 200 years.

Jean Thorey

1 RTE, Paris, FRANCE. jean.thorey @ rte-france.com

Introduction

We have several datasets of 200 years of hourly gridded data under future climate. Because power-flow simulations for 200 years are computationally costly, our objective is to find a small number of representative years.

Statistical models based on historical data convert weather variation to energy-related variables : consumption, and wind and photovoltaic (PV) capacity factors. Assumptions on the evolution of electricity generation and consumption (population growth, new wind power plants, technology evolution, electric mobility…) come from our Long-Term Adequacy Report [1].

The sample of roughly 10 years should be close to the full dataset. We resort to the definition of constraints to filter sample candidates.

Sample scoring

For each variable and location, we compute several scores between the sample mean and the full distribution:

- $\text{Energy score: } u\cdot Av - \frac{1}{2} u\cdot Av - \frac{1}{2} u\cdot Av$ Fast
- $\text{Quantile MAE: } \frac{1}{k}\sum_{s=1}^{N} |q(s) - \hat{q}(s)|$ Slow
- $\text{Quantile error: } |q(\hat{s}(a)) - \hat{q}(a)|$ Slow

Scores of random samples according to the sample size (number of years). The black line indicates the threshold $M_i$.

Table 1: Score definitions.

<table>
<thead>
<tr>
<th>Score</th>
<th>Formula</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean value</td>
<td>$</td>
<td>\mu - \mu' - \frac{1}{2} m - \frac{1}{2} m'</td>
</tr>
<tr>
<td>Energy score</td>
<td>$u\cdot Av - \frac{1}{2} u\cdot Av - \frac{1}{2} u\cdot Av$</td>
<td>Fast</td>
</tr>
<tr>
<td>Quantile MAE</td>
<td>$\frac{1}{k}\sum_{s=1}^{N}</td>
<td>q(s) - \hat{q}(s)</td>
</tr>
<tr>
<td>Quantile error</td>
<td>$</td>
<td>q(\hat{s}(a)) - \hat{q}(a)</td>
</tr>
</tbody>
</table>

Sampling strategies

We want $W_i, score_i < M_i$. Lowering $M_i$ increases difficulty.

To generate a new sample, we can use:

- random sampling
- random sampling based on a clustering of years [2]
- local search optimization.

To save computation time:
- we compute costly scores only if cheap scores are satisfying.

In practice, we tested $10^7$ samples for each experiment.

Experiment 1: weather

Data set 1 = weather-related:
- Climate projections 2050-2080 RCP 8.5.
- Weather variables : temperature, wind speed and solar radiation.
- 7 models $\times$ 210 years.
- 5 score types $\times$ 25 regional aggregates $\times$ 3 variables $\approx 375$ scores.

What about multi-model variability?
- We want to avoid overfitting, i.e. $M_i$ too low.
- We want $M_i >$ multi-model variability.
- With two samples $P_{90}, P_{90}$ of 90 years (2 $\times$ 3 models among 7), we set $M_i = \gamma \times \max_{s}(P_{90}, P_{90})$ with $\gamma = 1.5$.
- Numerical experiments for $M_i \approx 10^7$.

We illustrate:
- the large intermodel variability (large $M_i$).
- the importance of the sample size (number of years).
- the filtering efficiency of the number of constraints.
- the approximation error of a selected sample.

Experiment 2: energy

Data set 2 = energy-related:
- 200 years of Arpege-Climate model for 2050 under RCP 8.5.
- Weather $\rightarrow$ loads and renewables $\rightarrow$ power flows.
- 2 socio-economic scenarios.
- Energy variables: regional/national/continental net load, flows…
- Operational use of the sample of 10 years.

No other data? Optimize!
- Custom score thresholds $M_i$.
- Numerical experiments for $10^9 \approx 10^7$.

We illustrate:
- the difference between the sample and the full distribution.
- the approximation error of a satisfying sample.

Discussion

- Definitions of distance between years?
- Avoiding bad sampling vs. minimization?
- Number of scores?
- New variables with spatial or temporal aggregation?
- Definitions of score thresholds $M_i$?
- Sufficient number of climate models? Sufficient number of socio-economic scenarios?
- Sampling strategies?

Conclusion

- Lower approximation error with satisfying samples of years.
- Experiments on weather and energy data.
- A few R packages: fields, energy, lubridate, postorder.
- Code: https://github.com/rte-france/selectionsample

References